Gestion de données - notion : Statistiques

1. Vocabulaire

On appelle :

- **population** l'ensemble qui fait l'objet d'étude ;
- **individu**, un élément de la population ;
- et **caractère** ou **variable**, la propriété de la population qui est étudié.

Exemple :

Intéressons-nous à la note obtenue à un examen par chacun des étudiants d'un groupe dont **l'effectif total** est douze (n = 12) : 10 ; 13 ; 8 ; 14 ; 11 ; 8 ; 18 ; 10 ; 10 ; 14 ; 9 ; 16.

La population est l'ensemble des étudiants du groupe, un individu est un des étudiants du groupe, le caractère est la note obtenue à l'examen. En notant (x_i) où i varie de 1 à 12 cette **série** de notes, on a $x_1 = 10$; $x_2 = 13$; … ; $x_{12} = 16$.

Le caractère ou variable peut prendre plusieurs **modalités** (8, 9, 10, 11, 13, 14, 16 et 18 dans l'exemple ci-dessus).

Le caractère est dit **qualitatif** lorsque les modalités ne sont pas numériques, le caractère est dit **quantitatif discret** lorsque les modalités prennent un nombre fini de valeurs numériques, le caractère est dit **quantitatif continu** lorsque les modalités peuvent prendre un nombre infini de valeurs numériques ; dans ce cas, on regroupe souvent les modalités en intervalles ou classes.

Dans l'exemple précédent, le caractère est quantitatif discret.

2. Les caractéristiques de dispersion et de position

a) L'étendue

L'étendue d'une série quantitative est la différence entre la plus grande et la plus petite des modalités.

Dans l'exemple précédent, l'étendue est 10, elle s'obtient en calculant :

$\max(x_i) - \min(x_i) = 18 - 8 = 10$

L'étendue d'une série est une caractéristique de dispersion.

b) Le mode

Le **mode** ou la valeur (ou classe) modale d'une série est la modalité (ou classe) la plus fréquente.

Le mode est une caractéristique de position.

Dans notre exemple, le mode est égale à 10 (cette valeur apparaît trois fois, c'est plus que les autres valeurs).

c) La moyenne

La **moyenne** *m* d'une série statistique quantitative est le quotient de la somme de toutes les valeurs (prises par le caractère) par l'effectif total :

$$m = \frac{x_1 + x_2 + \dots x_n}{n}$$

Dans notre exemple :

$$m = \frac{10 + 13 + 8 + 14 + 11 + 8 + 18 + 10 + 10 + 14 + 9 + 16}{12} = 11{,}75$$

La moyenne est une caractéristique de position.

d) La médiane

La **médiane** M d'une série quantitative est un nombre tel que la moitié au moins des valeurs du caractère sont inférieures ou égales à M et la moitié au moins sont supérieures ou égales à M.

La médiane vise à partager la population en deux populations de même effectif.

Pour la déterminer, la première tâche consiste à ranger les valeurs x_i de la série par ordre croissant.

- Si l'effectif total *n* de la série est un nombre impair, c'est-à-dire qui peut s'écrire sous la forme 2k+1, alors la médiane de la série est la valeur centrale de la série : $M = x_{k+1}$.

- Lorsque l'effectif total *n* est pair de la forme 2k, la médiane n'est pas forcément une valeur de la série ; par convention, on choisit de prendre la demi-somme des deux valeurs centrales de la série ordonnée.

$$M = \frac{x_k + x_{k+1}}{2}$$

Dans notre exemple :

$$M = \frac{x_k + x_{k+1}}{2} = \frac{10 + 11}{2} = 10{,}5$$

La médiane est une caractéristique de position.

e) Les quartiles

Un nombre noté Q_1 est **premier quartile** d'une série si au moins un quart des valeurs de la série sont inférieures ou égales à Q_1 et si au moins trois quarts des valeurs de la série sont supérieures ou égales à Q_1.

Dans notre premier exemple :

8 ; 8 ; **9** ; **10** ; 10 ; 10 ; 11 ; 13 ; 14 ; 14 ; 16 ; 18

$Q_1 = 9{,}5$

Un nombre noté Q_3 est **troisième quartile** d'une série si au moins trois quarts des valeurs de la série sont inférieures ou égales à Q_3 et si au moins un quart des valeurs de la série sont supérieures ou égales à Q_3.

Dans notre premier exemple :

8 ; 8 ; 9 ; 10 ; 10 ; 10 ; 11 ; 13 ; **14** ; **14** ; 16 ; 18

$Q_1 = 14$

Les quartiles sont des caractéristiques de position.

La moitié des valeurs sont entre Q_1 et Q_3.

Dans notre exemple :

$Q_3 - Q_1 = 14 - 9,5 = 4,5$ est appelé l'intervalle interquartile.

f) Les fréquences

$$\text{Fréquence}_{\text{relative à une médiane}} = \frac{\text{effectif}_{\text{relatif à une modalité}}}{\text{effectif total}}$$

$$\text{Fréquence}_{\text{cumulée}} = \frac{\text{effectif}_{\text{cumulé}}}{\text{effectif total}}$$

$$\text{Taux}_{\text{exprimé en pourcentage}} = 100 \times \text{fréquence}$$